Course One
Foundations of Data Science

Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- ☑ ~~Complete the PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.~~
- ☑ ~~Create a project proposal for the data team.~~

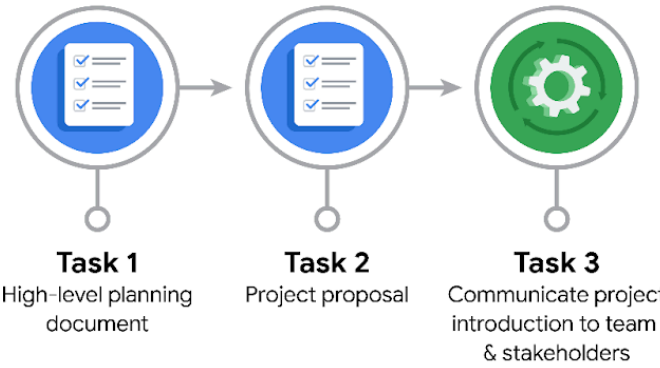
Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

- As a new member of a data analytics team, what steps could you take to get 'up to speed' with a current project? What steps would you take? Who would you like to meet with?
- How would you plan an analytics project?
- What steps would you take to translate a business question to an analytical solution?
- Why is actively managing data an important part of a data analytics team's responsibilities?
- What are some considerations you might need to be mindful of when reporting results?

Reference Guide

This project has three tasks; the following visual identifies how the stages of PACE are incorporated across those tasks.



Data Project Questions & Considerations

PACE: Plan Stage

- Who is your audience for this project?

The primary audience is the New York City Taxi and Limousine Commission (TLC). They are the client seeking to improve operational efficiency by predicting taxi fares. The Automatidata team is also an important audience, as they need to collaborate on the project tasks.

- What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

We aim to develop a regression model to predict taxi fares based on variables such as distance and time of day. The model will streamline fare estimates, improve accuracy for riders and drivers, and support the TLC's operational decision-making process.

- What questions need to be asked or answered?

  - What variables significantly affect fare predictions?
  - How accurate does the model need to be for real-world application?
  - What existing data should be used, and what additional data might be needed?

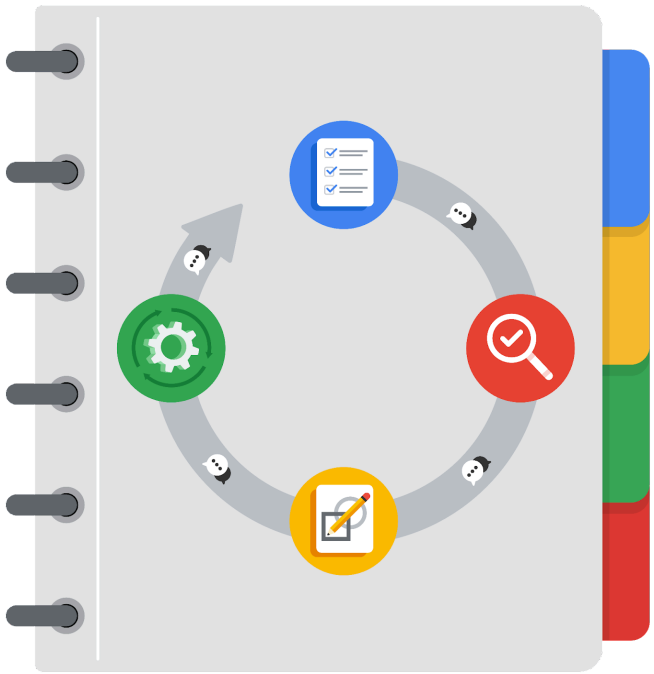
- What resources are required to complete this project?

Historical trip data (ride distance, time of day, fare details), software for regression modeling (e.g., Python, R), computational tools for running analyses, and data storage for managing large datasets.

- What are the deliverables that will need to be created over the course of this project?

A project proposal document, a regression model, a data summary, a report on model evaluation and testing, and a presentation of insights to TLC stakeholders.

THE PACE WORKFLOW



[Alt-text: The PACE Workflow with the four stages in a circle: plan, analyze, construct, and execute.]

You have been asked to demonstrate for the company's data team how you would use the PACE workflow to organize and classify tasks for the upcoming project. Select a PACE stage from the dropdown buttons. A few tasks involve more than one stage of the PACE workflow. Additionally, not every workplace scenario will require every task. Refer back to the Course 1 end-of-course portfolio project overview reading if you need more information about the tasks within the project.

Project tasks

Following are a group of tasks your company's data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop down menu. Next, give an explanation of why you selected the stage for each task. Review the following

readings to help guide your selections and explanation: The PACE stages and Communicate objectives with a project proposal. You will later reorder these tasks within a project proposal.

1. **Evaluating the model:** Analyze

   Why did you select this stage for this task?

A project proposal document, a regression model, a data summary, a report on model evaluation and testing, and a presentation of insights to TLC stakeholders.

2. **Conduct hypothesis testing:** Analyze **and** Construct

   Why did you select these stages for this task?

Testing hypotheses involves analyzing data to verify relationships and constructing tests to validate assumptions.

3. **Begin exploring the data:** Analyze

   Why did you select this stage for this task?

Exploration focuses on understanding the data's structure and initial insights before deeper analysis.

4. **Data exploration and cleaning:** Plan **and** Analyze

   Why did you select these stages for this task?

Planning how to clean the data and then executing the cleaning process to ensure data integrity.

5. **Establish structure for project workflow (PACE):** Plan

   Why did you select this stage for this task?

Structuring the project workflow involves defining the steps and resources before executing the project.

6. **Communicate final insights with stakeholders:** Execute

Why did you select this stage for this task?

Communicating insights is part of the execution phase, where results are shared with stakeholders.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Computing descriptive statistics helps in summarizing data and drawing initial conclusions.

8. **Visualization building:** Analyze **and** Construct

Why did you select these stages for this task?

Visualizations are created during analysis and construction phases to communicate findings effectively.

9. **Write a project proposal:** Plan

Why did you select this stage for this task?

Writing the proposal involves planning the structure and scope of the project.

10. **Build a regression model:** Construct **and** Analyze

Why did you select this stage for this task?

Building the model requires constructing it and analyzing the results to ensure accuracy.

11. **Compile summary information about the data:** Plan

Why did you select this stage for this task?

Summarizing the data is an early step to inform subsequent analysis.

12. **Build machine learning model:** Construct

Why did you select this stage for this task?

Constructing the model involves using machine learning techniques to build predictive models.